

# **InaNLP:** Indonesian Natural Language Processing Tools API



Ayu PurwariantiAguayu@informatika.orgagung.s	u <b>ng Sorlawan</b>	<b>; Sorlawan</b> Arief Rahman		Taufic Leonardo Sutedjo		
	sorlawan@prosa.ai	awan@prosa.ai arief.rahman@prosa.ai		taufic.sutedjo@prosa.ai		
<b>Sidik Soleman</b>	Dessi Po	<b>uji Lestari</b>	<b>Teguh E</b>	E <b>ko Budi</b>	<b>arto</b>	
sidik.soleman@prosa	n.ai dessipuji@in	nformatika.org	teguh.buc	diarto@pi	rosa.ai	
Introduction	InaNLP Com	ponents:		Tr	aining Data	
<ul> <li>InaNLP is an Indonesian Natural Language Processing Tools API</li> <li>Consists of several NLP tools:         <ol> <li>Lexical</li> <li>Syntactical</li> <li>Text classification</li> </ol> </li> </ul>	<ul> <li>POS Tagge</li> <li>Named end</li> <li>Dependend</li> <li>Constituer</li> <li>Word norm</li> <li>Quotation</li> </ul>	r tity tagger cy parser nt parser nalizer extraction		•	<ul> <li>The data annotation process was conducted by Indonesian linguists manually.</li> <li>Carefully process was conducted: <ol> <li>annotation guideline preparation</li> </ol> </li> <li>Data labelling</li> </ul>	

- Sentiment analysis Document level •
- Sentiment analysis Concept level ullet
- Topic classification. ٠

#### Part Of Speech Tagger

Easily integrated and combined

Built using deep learning algorithms

Training data	61,601 words			
Testing data	10,890 words			
Accuracy	96,13%			
26 POS Tag Classes	NNO, NNP, PRN, PRR, PRI, PRK, ADJ, VBI, VBT, VBP, VBL, VBE, ADV, ADK, NEG, CCN, CSN, PPO, INT, KUA, NUM, ART, PAR, UNS, \$\$\$, SYM			
Method	Bi-LSTM with word embedding, character embedding, prefix and suffix as the input layer			
Kalangan tokoh	mempertanyakan keputusan pengadilan mengenai kasus tersebut			

NNO	NNO	VBT	NNO	NNO	PPO	NNO	ART
prefix:	prefix:	prefix: <b>me, per</b>	prefix: <b>ke</b>	prefix: <b>pe</b>	prefix: <b>me</b>	prefix:	prefix: <b>te</b>
suffix:	suffix:	suffix:	suffix:	suffix:	suffix:	suffix:	suffix:
- derivational: <b>an</b>	- derivational:	- derivational: <b>kan</b>	- derivational: <b>an</b>	- derivational: <b>an</b>	- derivational: i	- derivational:	- derivational:
- particle :	- particle :	- particle :	- particle :	- particle :	- particle :	- particle :	- particle :
- possesive :	- possesive :	- possesive :	- possesive :	- possesive :	- possesive :	- possesive :	- possesive :

menyukai

VBT

saya

PRN

## Named Entity Tagger

Training data	268,522 words
Testing data	36,680 words
Accuracy	92,1%
NE Tag classes are the Ontonotes 5 with 18 classes	PER, NOR, FAC, ORG, GPE, LOC, PRO, EVT, WOA, LAW, LNG, DTE, TME, PCT, MON, QUA, ORD, CRD
Method	Bi-LSTM with word embedding and character embedding

3. Quality checking.

NOR	PERSON	FACILITY	DATE	GPE	EVENT
-----	--------	----------	------	-----	-------

Presiden Indonesia, Joko Widodo bertemu dengan Presiden Amerika Serikat Barrack Obama di Istana Bogor pada hari Kamis. Setelah itu, Jokowi dan Obama bersama-sama menuju <mark>Jakarta</mark> untuk menyaksikan <mark>Asian Games 2018</mark>.

### **Dependency Parser**

Training data	1440 sentences	Training da
Testing data	400 sentences	Testing da
Performance score	UAS=80.6% and LAS=70.3%	Performar
There are 34 relation types (modified from Universal Dependencies V2 conventions)	acl, acl:relcl, advcl, advmod, amod, appos, aux, case, cc, ccomp, clf, compound, conj, cop, csubj, det, discourse, fixed, flat, goeswith, iobj, list, mark, nmod, nmod:poss, nsubj, nummod, obj, obl, orphan, parataxis, punct, vocative, xcomp	Method
Method	Shift reduce parser using MLP with features of word embedding and result of Bi-LSTM on character embedding	
nsubj	obj compound	

bahasa

NNO

## **Word Normalization**

Training data	569,982 words
Testing data	52,692 words
Performance score	F1 score = 90.1%
Method	divided into detection and correction. In detection module, we use bi-lstm with word embedding,

character embedding and orthography information to detect which word should be normalized. In correction module, we choose words from dictionary with smallest word distance.

Kmrn liat foto2 Bdg jaman doeloe, kereeennn



Indonesia

NNP



Kemarin liat foto-foto bandung zaman dahulu, kereeennn